

Bc. Jana Kravalová: Využití syntaxe v metodách vyhledávání informací

Posudek vedoucího diplomové práce

Diplomová práce Jany Kravalové nese název *Využití syntaxe v metodách vyhledávání informací* a má výzkumně-experimentální charakter. Jejím cílem je zapojení syntaktické informace do metod vyhledávání informací založených na jazykovém modelování. Mimo to studuje i další techniky používané v této oblasti a jejich vliv na úspěšnost vyhledávání. Je třeba poznamenat, že v oboru vyhledávání informací v češtině je to práce do značné míry pilotní – jedná se o první práci, která vyhodnocuje a srovnává úspěšnost různých metod pro vyhledávání informací na doposud jediné české testovací kolekci (vytvořené v rámci *Cross-Language Evaluation Forum* v roce 2007).

Předkládaná práce je co do rozsahu poměrně krátká (je tvořena šesti kapitolami na celkem 44 stranách), ale co do obsahu je poměrně bohatá a předkládá čtenáři řadu zajímavých zjištění.

První kapitola obsahuje úvod a motivaci pro celou práci, včetně specifikace jejích cílů. Hlavní cíl spočívá v navržení, implementaci a vyhodnocení modelů pro vyhledávání informací, které by využívaly syntaktickou informaci získanou automaticky jak z dokumentů, tak dotazů výše zmíněné kolekce. Další cíle zahrnují jejich srovnání s tradičními modely a vyhodnocení přínosu standardních technik (včetně jejich implementace), jako např. použití lemmatizace, stemmingu, vyhlazování, vážení slov dotazů a pseudo-relevance feedbacku. Dále se autorka pokouší několika způsoby kombinovat různé modely a zvýšit tak celkovou úspěšnost vyhledávání. V druhé kapitole diplomantka seznamuje čtenáře s teorií vyhledávání informací (zejména s principem použití jazykových modelů) a popisuje použité modely (unigramový, povrchový bigramový a syntaktický bigramový) a jejich varianty. Pozornost je také věnována způsobu evaluace, která je založena na míře Average Precision. Ve třetí kapitole jsou popsána data, na kterých byly prováděny experimenty, a jejich rozdělení na trénovací a testovací část. Jádrem práce jsou kapitoly 4 a 5 obsahující popis použité metodologie, detaily experimentů (např. způsob kombinace modelů) a prezentaci výsledků. Kapitola 6 obsahuje závěr celé práce.

Pro provedení experimentů bylo nutné vytvořit software středního rozsahu; nástroje pro parsing, značkování a lemmatizaci byly převzaty.

Hodnocení

Předkládaná práce plně splňuje požadavky kladené na diplomovou práci na MFF UK. Všechny její cíle byly splněny. Práce vyniká zejména v rozsahu provedených experimentů, systematickosti a korektnosti postupu při jejich provádění. Textu by bylo možné vytknout snad jen přílišnou stručnost v některých pasážích stěžující jejich pochopení (např. poslední odstavec v části 2.4.2) nebo naopak příliš detailní a méně přehledné obrázky a tabulky (např. obr. 5). Celkově je práce ovšem nadprůměrná a doporučuji její výsledky publikovat na vhodné konferenci či workshopu. Diplomovou práci doporučuji k obhajobě.